# SYSTEM FOR RETRIEVING INFORMATION FROM A PLURALITY OF REMOTE DATABASES HAVING AT LEAST TWO DIFFERENT LANGUAGES

## BACKGROUND OF THE INVENTION

This invention relates to the field of information retrieval. More specifically, the invention comprises an apparatus and method for obtaining information from any one of the multiplicity of commercially available databases. The invention makes such information retrieval simple, efficient, and economical.

It has long been recognized that information of any kind, including stored data, can be transmitted by telecommunications technology. One example of an information retrieval system for providing data to a plurality of subscribers is shown in U.S. Pat. No. 4,450,477. Other examples of systems for transmitting data to subscribers in remote locations are given in U.S. Pat. Nos. 4,337,485, 4,303,941, 3,751,670, and 4,395,780. The main disadvantage of the system shown in U.S. Pat. No. 4,450,477 is that it requires a complete cable television system for its operation. While all of the cited patents deal with storage and transmission of data, none addresses the special problems encountered in dealing with a multiplicity of commercially available databases.

There are hundreds, if not thousands, of commercial databases now in existence. As used herein, the term "database", or "commercial database", means a collection of data, usually concerning one particular field or subject, which is stored in the memory of a computer (or on peripheral storage devices), and which is accessible to a plurality of users, the users being situated in remote locations. The users establish contact with the databases through telecommunication lines. Normally, the users pay a fee for the service, the fee being based on the amount of time spent in communication with the database.

Until comparatively recently, in order to have access to a commercial database, the user needed a special terminal, which generally was fairly expensive. In recent years, with the proliferation of personal computers in homes and offices, the number of persons having equipment capable of communicating with outside databases has increased dramatically. Now, anyone with a personal computer, a telephone, and a modem (a device for converting the digital signals of the computer into analog signals capable of being transmitted over telephone lines, and vice versa) can gain access to any commercial database. The huge potential market for such information services has contributed to the dramatic increase in the number of commercial databases.

Unfortunately, each commercial database has a "language" of its own. That is, for each database, there is a different set of syntactical rules governing the formulation of search requests. A user who desires to search in a large number of databases must learn all the different languages for the databases. Although different databases often have languages which are similar in some ways, the differences that remain, subtle as they may be, cannot be ignored. The user must therefore maintain a working knowledge of each such language. If the user is not thoroughly familiar with the database language before establishing communication with that database, the user may waste valuable time in attempting to formulate a search request correctly. Such a procedure is obviously an expensive one, as the fee for using commercial databases is linked to the time spent "on-line" with the database.

The expense incurred in using a commercial database can be divided, in general, into two components. There is a charge made for the "search" time, which is time spent actually searching the database for documents fulfilling the search request. And there is a charge for "browse" time, which is the time spent in reviewing the documents retrieved by the search. Although some database vendors may charge less for browse time than for search time, the user almost invariably requires much more browse time than search time. Because the user must maintain communication with the database in order to browse, the cost of even a straightforward search can become prohibitive.

Another basic problem resulting from the large number of commercial databases is that many potential users do not know what databases are available. Unless one is a skilled reference librarian in the ever-changing world of commercial databases, one is unlikely to be aware of all the databases which may contain the solution to a researach problem. This problem affects not only the person who is inexperienced in the use of commercial databases, but also the expert user whose knowledge of new databases may not be current.

The present invention solves all of the problems described above, by providing a new apparatus and method for communicating with a large number of commercial databases. The invention eliminates the need for the user to know what database should be searched. The invention also makes it unnecessary to learn more than one language for formulating search requests. The searches performed with the system of the present invention are, on average, considerably less costly than searches done by conventional methods.

The present invention also helps those persons who are already expert in the use of commercial databases, by reducing the number of commands necessary to gain access to a wide variety of databases. With the present invention, the even the expert user can save valuable time in performing searches.

## SUMMARY OF THE INVENTION

The present invention, in its simplest form, comprises one or more programmed digital computers connected, by suitable modems, to incoming telephone lines, from users, and to outgoing telephone lines, to various databases. The user dials the telephone number of the system, which, after asking for the user's credit card number, presents the user with a series of questions concerning the area in which the user wishes to search. The questions are pre-programmed, according to a predetermined decision "tree", and are designed to provide the user with choices that span virtually the entire range of human knowledge. After the user responds to a given question, the systems asks another question, in order to determine more precisely what database(s) would be most appropriate to solve the user's problem.

Based on the user's responses to the system's inquiries, the system chooses a commercial database that is believed to be most likely to contain the solution to the user's research problem. The computer then asks the user for a search request. A search request comprises a word, or a group of words connected by logical operators. The object of the search is to retrieve all documents in the database which contain that word or group of words. The system requires the user to be familiar